Examining Relationships Checkpoint 2

Question (1)

High blood pressure is unhealthy. Here are the results of one of the studies that link high blood pressure to death from cardiovascular disease. The researchers classified a group of white males aged 35 to 64 as having Low blood pressure or High, then followed the subjects for five years. The following two-way table gives the results of the study:

	Blood pressure		
Cardiovascular death?	Low	High	Total
Yes	21	55	76
No	2655	3283	5938
Total	2676	3338	6014

In this example, it would be appropriate to calculate:

A: conditional row percentages

B: conditional column percentages

C: the correlation coefficient r

D: the five-number summary of both variables

E: none of the above

Feedback

A : 0

✗ This is not quite right. Remember that to explore a categorical relationship, the most appropriate summary is the outcome percentages within each ***explanatory*** group. For the table shown, think about whether the row variable (cardiovascular death) or the column variable (blood pressure) would be the ***explanatory*** variable. (b) is the correct answer.

B : 10

✓ Good job! To explore a categorical relationship, the most

✔ appropriate summary is the outcome percentages within each ***explanatory*** group. In this case, the column variable (blood pressure) would be the explanatory variable, so the appropriate percentages are the death percentages from each blood pressure group.

C : 0

✘ This is not quite right. Remember that the correlation coefficient is a summary value for a scatterplot, not for a ***two-way table***. Try to remember the most appropriate summary for exploring a ***categorical*** relationship. (b) is the correct answer.

D : 0

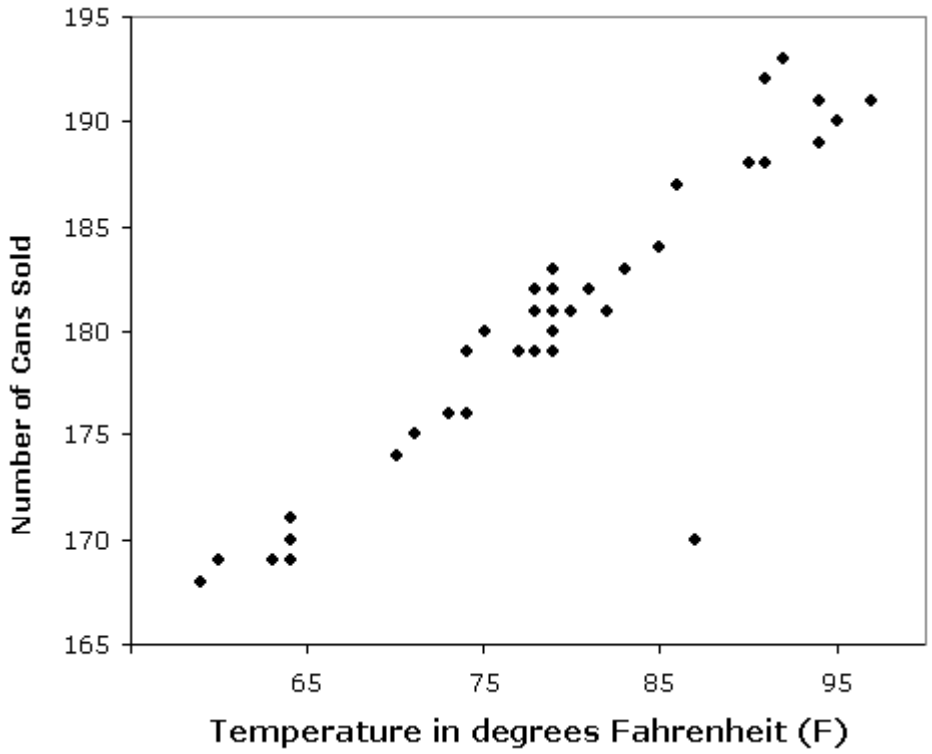
✘ This is not quite right. Remember that five-number summaries are only appropriate for lists of ***quantitative*** measurements. But a two-way table summarizes ***categorical*** variables, not quantitative variables. Try to remember the most appropriate summary for exploring a ***categorical*** relationship. (b) is the correct answer.

E : 0

✘ This is not quite right. Notice that the two-way table summarizes categorical variables. Try to remember the most appropriate summary for exploring a ***categorical*** relationship. (b) is the correct answer.

Question (2)

A local ice cream shop kept track of the number of cans of cold soda it sold each day, and the temperature that day, for two months during the summer. The data are displayed in the scatterplot below:



The one outlier corresponds to a day on which the refrigerator for the soda was broken. Which of the following is true?

A: A reasonable value of the correlation coefficient r for these data is 1.2.

B: If the temperature were measured in degrees Celsius (C = 5/9*(F-32)), the value of r would change accordingly.

C: If the outlier were removed, r would increase.

D: If the outlier were removed, r would decrease.

E: Both (b) and (c) are correct.

Feedback

A : 0

✗ This is not right. Remember that the correlation coefficient r must be between â€“1 and 1. It cannot be larger than 1. Think about the other choices. (c) is the correct answer.

B : 0

✗ This is not quite right. Remember that the correlation coefficient r is a unitless number, so it's not affected by a change of units. Think about the other choices. (c) is the correct answer.

C : 10

✓ Good job! The outlier makes the data less "tightly clustered"

✓ around the linear trend. Removing the outlier will increase the overall tightness of the points around their straight-line trend, so the correlation coefficient would be closer to 1.

D : 0

✗ This is not quite right. Remember, for a scatterplot where the trend is generally rising, the ***more*** linear the scatterplot is, the ***closer to 1*** the correlation coefficient r will be. Notice that removing the outlier will ***increase*** the overall tightness of the points around their straight-line trend. (c) is the correct answer.

E : 0

✗ This is not quite right. It's true that the correlation coefficient would get closer to 1. But remember that the correlation coefficient r is a ***unitless*** number, so it's not affected by change of units. (c) is the correct answer.

Question (3)

Suppose that the correlation r between two quantitative variables was found to be r = 0. This means that:

A: there is a strong linear relationship between the two variables.

B: there is no linear relationship between the two variables.

C: there is a strong relationship between the two variables.

D: there is no relationship between the two variables.

E: none of the above.

Feedback

A : 0

✗ This is not quite right. Remember that a correlation coefficient ***far*** from zero would indicate a strong linear relationship. Think about the other choices. (b) is the correct answer.

B : 10

✓ Good job! The correlation coefficient indicates the strength of

✓ the linearity of a scatterplot relationship, and a correlation coefficient that is zero indicates no linear relationship. Remember, however, that there might still be some other (non-straight-line) kind of relationship.

C : 0

✗ This is not quite right. Remember that a correlation coefficient ***far*** from zero would indicate a strong linear relationship. Also, notice this choice didn't specify "linear." Think about the other choices. (b) is the correct answer.

D : 0

✗ Almost! But read this choice carefully. Remember that the correlation coefficient only measures ***linear*** relationship. (b) is the correct answer.

E : 0

✗ This is not quite right. Remember that the correlation coefficient indicates the ***strength of the linearity*** of a scatterplot relationship. Think about the other choices. (b) is the correct answer.

Question (4)

A correlation of r = .85 is found between weekly sales of firewood and cough drops over a 1-year period. Which of the following is true?

A: There is a pretty strong positive linear relationship between sales of firewood and cough drops.

B: Fire must be the cause of coughing.

C: Temperature is a possible lurking variable that is “behind” this relationship.

D: Both (a) and (c) are true.

E: None of the above.

Feedback

A : 0

✗ Almost! But remember that the correlation coefficient by itself isn't enough to determine whether the relationship was linear. We'd need to see the scatterplot to determine if the relationship was a straight-line trend. Think about the other choices. (c) is the correct answer.

B : 0

✗ This is not quite right. Remember that we can't determine ***causation*** merely from the nature of a display. Just because there might be an ***association***, that doesn't (by itself) indicate anything about ***causation***. Causation can only be determined by knowing the nature of the study design. (c) is the correct answer.

C : 10

✓ Good job! The sales of both items will tend to be low in the warm parts of the year, and tend to be high in the cold parts of the year. So the sales would tend to be positively associated (both low at the same time, both high at the same time), which would explain the strong positive correlation coefficient.

D : 0

✗ Almost! But remember that the correlation coefficient by itself isn't enough to determine whether the relationship was linear. We'd need to see the scatterplot to determine if the relationship was a straight-line trend. (c) is the correct answer.

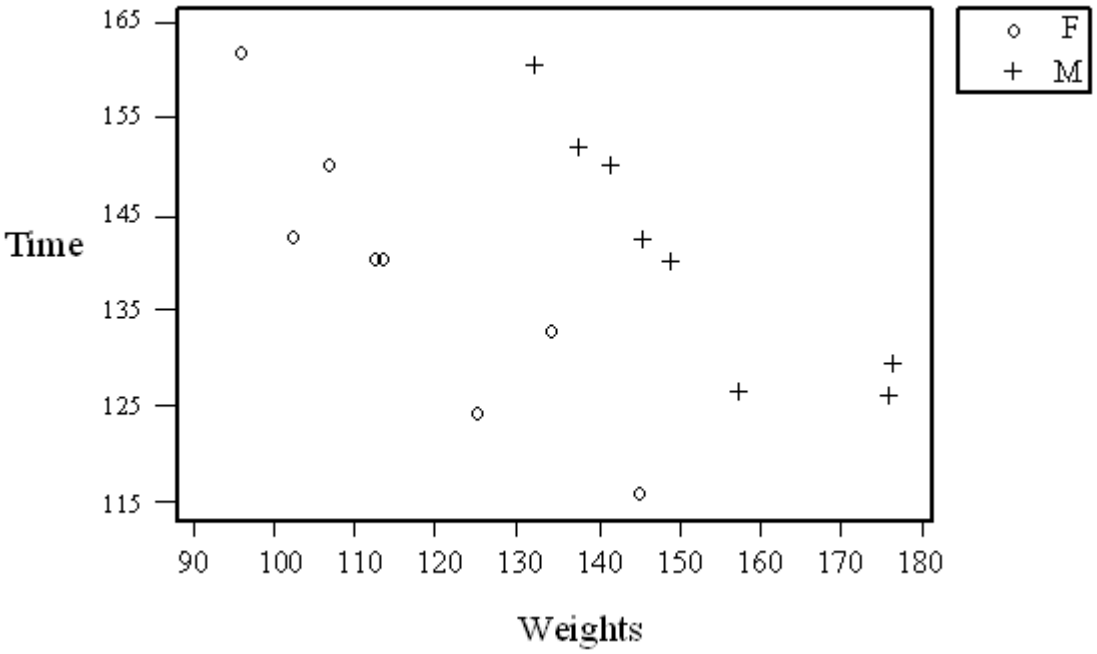
E : 0

✗ This is not quite right. Remember that a positive correlation coefficient would suggest a positive relationship (so when X was low, Y would be low, and when X was high, Y would be high). Think about what might reasonably explain that association in this case. (c) is the correct answer.

Question (5)

The data in the scatterplot below are an individual's weight and the time it takes (in seconds) on a treadmill to raise his or her pulse rate

to 140 beats per minute. The o's correspond to females and the +'s to males. Which of the following conclusions is most accurate?



A: There is a positive correlation between gender and weight, since men tend to weigh more.

B: There is a negative correlation between time and weight for males and for females.

C: There is a positive correlation between time and weight for males and for females.

D: Both (a) and (b) are correct.

E: Both (a) and (c) are correct.

Feedback

A : 0

✗ This not quite right. Remember that there can only be a "correlation" between two ***quantitative*** variables. But notice that "gender" is categorical. Reconsider the other choices. (b) is the correct answer.

B : 10

✓ Good job! For each separate gender, we see that the points on the scatterplot tend to ***fall*** from left to right. This indicates a ***negative*** relationship. So the correlation coefficient would be negative for males, and would also be negative for females.

C : 0

✗ This is not quite right. Remember that the correlation coefficient

would be ***positive*** only when the points tend to ***rise*** from left to right. But on the scatterplot, we see that, for both males and females, the points tend to ***fall*** from left to right. (b) is the correct answer.

D : 0

Almost! It's true that there is a negative correlation between time and weight for each separate gender, since we see that the points fall from left to right. But remember that there can only be a "correlation" between two ***quantitative*** variables. Notice that "gender" is categorical. (b) is the correct answer.

E : 0

This is not quite right. Notice that "gender" is categorical. Remember that there can only be a "correlation" between two ***quantitative*** variables. Next, notice that on the scatterplot we see that, for both males and females, the points tend to ***fall*** from left to right. Remember that the correlation coefficient would be ***positive*** only when the points tend to ***rise*** from left to right. (b) is the correct answer.

Question (6)

What can we say about the relationship between the correlation r and the slope b of the least-squares line for the same set of data?

A: Both r and b always have values between -1 and 1.

B: r is always larger than b.

C: r and b have the same sign (+ or -).

D: the slope b is always equal to the square of the correlation r.

E: b is always larger than r.

Feedback

A : 0

This is not quite right. It's true that the correlation r always has a value between â€“1 and 1. But there is no such restriction on slope. The slope of a line can exceed 1 or â€“1. (c) is the correct answer.

B : 0

✗ This is not quite right. Remember that the correlation r always has value between â€“1 and 1. But a regression line could have a slope larger than 1. So it's possible for b to be larger than r. (c) is the correct answer.

C : 10

✓ Good job! Although the correlation r isn't the same as the slope b, the thing they always have in common is their ***sign***.

D : 0

✗ This is not quite right. There is no such relationship between slope and correlation. Remember that the correlation r always has a value between â€“1 and 1, so the square of the correlation will also be between â€“1 and 1. But there is no such restriction on slope. (c) is the correct answer.

E : 0

✗ This is not quite right. Remember that for points that tend to rise from left to right, the correlation is close to 1 when the points are tightly clustered around the regression line. But if the regression line is shallow, the slope could be close to zero. So it's possible for b to be ***less*** than r. (c) is the correct answer.

Question (7)

A study was done on the timeliness of flights (on-time vs. delayed) of two major airlines: StatsAir and AirMedian. Data were collected over a period of time from five major cities and it was found that StatsAir does better overall (i.e., has a smaller percentage of delayed flights). However, in each of the five cities separately, AirMedian does better.

Which of the following is correct?

A: This situation is mathematically impossible.

B: This is an example of Simpson's Paradox.

C: "City" is a lurking variable in this example.

D: This is an example of a negative association between variables.

E: Both (b) and (c) are correct.

Feedback

A : 0

✗ This is not quite correct. Remember that it sometimes ***can*** happen that there is an association that is seen a particular way for each of several groups separately, but which can "reverse" when the groups are combined. (e) is the correct answer.

B : 0

✗ Almost! It's true that this is an example of Simpson's Paradox, because there is an association that is seen a particular way for each of several groups separately, but which can "reverse" when the groups are combined. But remember that such reversal typically indicates a lurking variable. (e) is the correct answer.

C : 0

✗ Almost! It's true that 'city' is a lurking variable, because there is an association that is seen a particular way for each of several groups separately, but which "reverses" when the groups are combined. But remember that such reversal for categorical variables has a particular name. (e) is the correct answer.

D : 0

✗ This is not quite right. Remember that it is only appropriate to call an association "positive" or "negative" when the variables are ***quantitative***. But in this case, both variables ("timeliness" and "city") are categorical. (e) is the correct answer.

E : 10

✓ Good job! This is an example of Simpson's Paradox, because there is an association that is seen a particular way for each of several groups separately, but which "reverses" when the groups are combined. Such reversal typically indicates a lurking variable, which in this case is "city."

Please answer the question below. Your response will not be graded, but it will be available for your instructor to read.

When faced with a data analysis problem that involves two variables, explain how you would decide which graphical display and numerical measures to use.

Feedback

*** : 0**

✗ Thank you for your response. It has been recorded and will be available for your instructor to read.

✗ You did not enter a response for this question.